Mobile Health

Dr Robert Harle
Mobile and Sensor Systems
Lent 2020

Definition

Providing healthcare support, delivery and intervention via mobile devices such as smartphones and wearables

Mobile Health Evolution

In 2009, UN Foundation offered a taxonomy of mobile health applications::

- Education and awareness
- Helpline
- Diagnostic and treatment support
- Communication and training for healthcare workers
- Disease and epidemic outbreak tracking
- Remote monitoring
- Remote data collection

Back in 2009 the focus was very much on the devices providing communication channels and trend analysis from the large penetration of mobiles.

Mobile Health Evolution

Since then mobile devices have become more powerful, more pervasive, more portable/wearable and packed full of sensors.

The fastest growing subset of mobile health exploits these changes, and that will be the main focus of this course since it continues the sensing theme.

Mobile Health Sensing

The Key Shift in Sensing

Mobiles offer **lower fidelity** sensing at much **higher availability** than clinical sensing

The Key Shift in Sensing

Mobiles offer **lower fidelity** sensing at much **higher availability** than clinical sensing

Typically the sensors are chosen to keep cost and size down even when this compromises on accuracy.

Measurements are not typically taken in controlled conditions as they would be in the clinic, increasing noise and error.

Instead of just getting a one-off measurement in a clinical setting, we have huge numbers of measurements taken in an uncontrolled real-world situation

("**ambulatory sensing**")

The Key Shift in Sensing

Mobiles offer **lower fidelity** sensing at much **higher availability** than clinical sensing

We are only in the very early stages of understanding what this means for health sensing

Types of Mobile Health Sensing

Applying clinical sensors and techniques directly

Taking established clinical sensors and applying them to mobiles esp. wearables

E.g. PPG heart rate, Blood pressure cuffs

Applying different sensors to emulate clinical techniques

Taking a clinical test and using available mobile sensors to do something similar.

E.g. Gait tests measure walking sway and symmetry using special sensors. We can approximate them using inertial sensors.

Completely new techniques

Exploiting new measurements we have never had before to develop new medical science.

E.g. 24/7 heart rate variability measures.

The Validation Challenge

Whatever path we take, we face difficulties in showing that our tests or outputs from our sensing are correct.

New *clinical* sensors/algorithms are simpler - you have a gold standard and you show how well you do against it. You get lots of people into the clinic and take the measurements.

Going mobile and adding ambulatory sensing imposes a major issue: the gold standard is very rarely available 24/7 so we have lots of measurements and results without ground truth!

The Validation Challenge: Example

Passive blood pressure measurement is a current 'holy grail' for wearables and there are various companies claiming to do it.

But how would you know it's working? You can certainly compare to a blood pressure cuff in a clinical setting and look at accuracy stats.

But just because it works there doesn't mean it's working in its actual use case, outside the clinic...

Case study: PPG

Photoplethysmography (PPG)

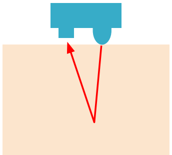
Detects blood volume changes and the standard way to measure heartrate and blood oxygen saturation in clinical/medical environments.

Simple combination of a bright light and a photodiode, Most are transmissive, meaning the light passes through the finger/earlobe to the photodiode

Technology Applied to (Smart/Sports) Watches

Taking a simple technology and applying it to the watch introduces a number of compromises:

- Transmissive doesn't work because light can't penetrate enough. So we use reflective. Much weaker signal, more affected by skin tone
- Minimise power draw so as to preserve watch battery
- Reduce sampling rate to preserve watch battery



Lower Fidelity: Uncontrolled Conditions

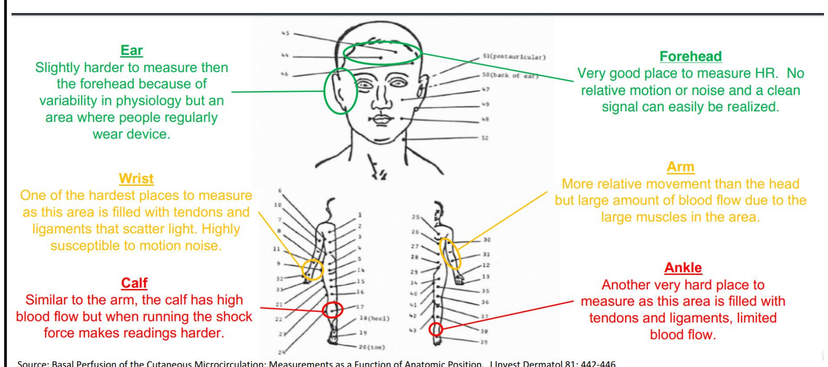
Clinical PPG has an excellent quality, mainly because measurements are taken under ideal conditions: ideal measurement site, no movement, no interference, properly worn.

Watches are different:

- The wrist is not an ideal measurement site for *anything!*
- Movement is commonplace.
- External light can get in.
- Watches may be worn loose.

Note that, under clinical conditions, the watch PPG does very well. The bonus is we get measurements all day. The quid-pro-quo is that most (all?) are not in clinical conditions...

Where to measure PPG

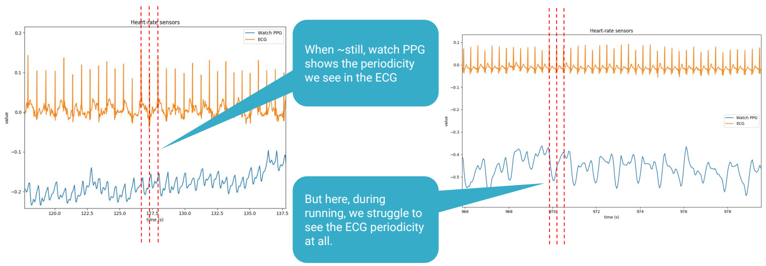


Source: Basal Perfusion of the Cutaneous Microcirculation: Measurements as a Function of Anatomic Position, J Invest Dermatol 81: 442-446

How Bad is It?

All the sinew, tendon and bone in the wrist means the SNR for the heart rate is very low: it doesn't take much to drown out the HR signal

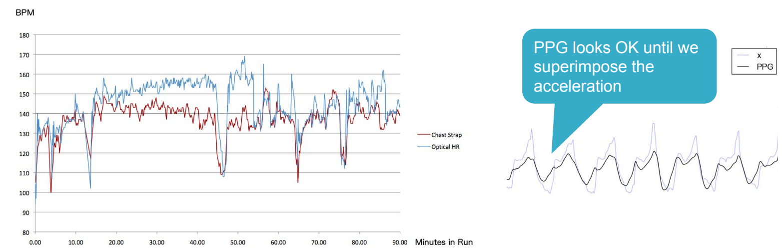
Rapid movements results in very poor PPG



Cadence Lock

Worse. If you are generating a strong repetitive signal in the PPG at a heartrate-like frequency the HR algorithms will lock onto that!

This commonly occurs in running, where it goes by the name of **cadence lock**. Although a lot of runners don't realise they aren't looking at their heartrate!



Even so...

Having HR available in-situ all day for millions is allowing us to explore new science:

- Daily HR distribution
- Nighttime HR and sleep analysis
- Resting heart rate when awake and asleep
- Effect of illness, epidemic tracking and recovery times
- Etc.

Although, again, validation is a challenge!

Mobile Health Sensing Applications

Mobile Health Applications

Remote diagnostics *

Simplifying access to health data *

Screening for conditions *

Monitoring chronic conditions

Detecting adverse events *

Participatory sensing (crowdsourcing) *

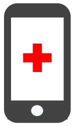
Supporting behaviour change

Education about and interpretation of public health guidelines *

We will look in more detail at those with an asterix

Remote Diagnostics

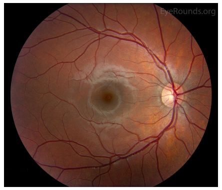
The Phone as the Clinical Device



In many parts of the world, access to specialist medical personnel and equipment is very lacking.

However, mobile phones are almost as ubiquitous as here and the sensors can be adapted to take and transmit measurements that allow for a remote expert to make a diagnosis

Example: Smartphone Funduscopy



Eye diseases are a major medical issue in developing nations.

A clinical fundus camera (retinal imager) costs £1000s and is out of reach of towns where it is needed most.

MIT and others have shown you can use a smartphone camera with an inexpensive lensing system to take retinal photographs for local or remote diagnosis.

Example: SpiroSmart

University of Washington.

Measures lung function by having you blow hard into the phone microphone.

A great example of repurposing the phone's sensors.

Simplifying Access to Health Data

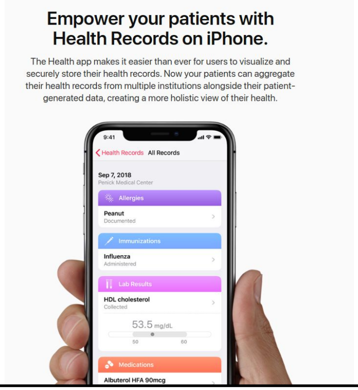
DeepMind Streams

Sometimes the biggest impact is from the simplest things.

Streams is an app for Doctors that collates patient info into one place. It proactively alerts Doctors when test results come back that aren't good.

Currently on trial at the Royal Free Hospital in London, where it is being used to relay test results for Acute Kidney Injury.

Apple Health Records



Screening

Screening

Screening for a disease is to test for it in someone not known to have that disease.

A positive result leads to further investigation and (if the screening test was correct, which it isn't always) a medical diagnosis.

The goal is to catch diseases early, before serious symptoms have developed and while it is (generally) easier to treat.

Screening on Mobiles/Wearables

A screening test that can be administered on (ideally unmodified) mobile devices:

- Is cheaper to carry out;
- Has the potential to reach many more people and remote places;
- Reduces the self-selection aspect (the sort of person who responds to screening calls is often not the type of person you need to target);
- Can be based on longitudinal data. Most screening programmes today involve quick tests in a lab environment. Wearables offer always-on longitudinal data for different/further assessment.

Sensitivity and Specificity

Medics evaluate a test using Sensitivity and Specificity

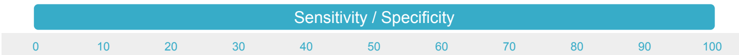
Sensitivity. True positive rate: of all actual positives, how many did the test find? TP/(TP+FN). *In CS/ML this is often called recall*.

Specificity. True negative rate: of all actual negatives, how many did the test find? TN/(TN+FP).

What about *precision*? Computed as TP/(TP+FP), it measures how likely a positive test is to be right, and that is surely valuable. Actually, Medics have a different name for it: Positive Predictive Value (PPV) and you need to be careful with it...

Question

You are being screened for failitis, a serious condition where you fail all exams. The test comes back positive! Assuming the test has equal sensitivity and specificity, when do you start getting really concerned that you have it?



Question

The Apple watch can detect Atrial Fibrillation with sensitivity 0.98 and specificity 0.996

You start wearing one and it alerts you to Atrial Fibrillation. How likely are you to have AFib?

Intuitive View

AFib Prevalence

The chance that a positive test is correct is the **Positive Predictive Value (PPV)** to medics and the **Precision** to you,

PPV = Precision = TP / (TP+FP)

The problem is that this value depends strongly on the **prevalence**: how common a positive is in the 'wild'.

AFib Prevalence

The chance that a positive test is correct is the **Positive Predictive Value (PPV)** to medics and the **Precision** to you,

PPV = Precision = TP / (TP+FP)

The problem is that this value depends strongly on the **prevalence**: how common a positive is in the 'wild'

This makes sense: if something is very unlikely to occur, then taking a sample of people to test on will give very few true positives (because it's so rare) and a lot of false negatives (because there are a lot of chances for them). So the PPV/precision is low.

The initial non-intuitive result is called the **base rate fallacy**.

Apple AFib PPV

Consequences for Mobile Device Screening

Even if you have a 'good' test, if the prevalence of the disease is very low (as with many diseases) you will have a low precision and generate many false positives.

This, in turn, means health services overrun with people worried they have a disease when in fact they don't. Paradoxically this takes care away from those who do have it!

The nature of consumer mobiles is that the sensors are not going to be clinical grade. So the sensitivities we need to screen for very rare diseases are often unachievable.

Lesson 1: You'll get the best return for more prevalent diseases.

Lesson 2: The lower the prevalence the higher you need to make the specificity.

Important Note

Wearables generate large volumes of data and creating an ML classifier is a sensible approach.

Often you want to balance each class in training. You might go out and find a clinic with 100 patients with the disease, and then collect data from 100 healthy controls. This is a reasonable way to develop your classifier.

However, when it comes to evaluating its value as a screening tool, remember to incorporate the prevalence and not just compute the precision from your collected data. This is a common error.

Detecting Adverse Events

Example: Fall Detection

Playing to the strengths of wearables: accelerometers always on

Falls characterised by a short period of free fall (accelerometer measures zero on all axes) followed by a sharp impact and (possibly) a period of no movement

You might like to think how to validate this one…

(To see how Apple did it, visit https://www.popsci.com/apple-watch-fall-detection/)

Coughs and Sneezes

Coughs and sneezes have strong audio characteristics that makes them amenable to low-cost recognition. A number of papers have discussed the possibility of always-on cough detection on mobile devices.

There are three obvious things we might want to do here:

- Count the number of coughs to understand whether a respiratory problem is getting better or worse.
- Classify the type of cough to decide whether to prioritise getting treatment.
- Track the progression of a cold/flu throughout the population

Participatory Sensing

Participatory Sensing

The wide penetration of mobile devices makes them an ideal sensing network: users carry them everywhere and take responsibility for charging and maintaining them.

Participatory sensing is the use of a crowd to map some quantity.

It can allow mapping at a scale and granularity that would not otherwise be feasible.

Example: Pollution Mapping

WHO: Air pollution kills 7,000,000 people annually.

Can now measure gases using tiny devices integrated into wearables.

E.g. Ieva: a smart keyfob that measures nitrogen dioxide, ozone, sulphur dioxide, fine particles and contributes to a global mapping.

Interpretation of Public Health Guidelines

Interpreting Guidelines

Public health guidelines should to be simple to understand and follow.

But the science is often complex and keeping the message simple can make it difficult for an individual to apply it.

Mobile health offers an opportunity to **personalise guidelines**, taking into account the nuances in the science, making them more relevant and hence more likely to be followed.

Case Study: Physical Activity Guidelines

Here's the activity guideline for cardiovascular fitness adopted worldwide:

"Adults should do at least 150 minutes of moderate-intensity aerobic physical activity throughout the week or do at least 75 minutes of vigorous-intensity aerobic physical activity throughout the week or an equivalent combination of moderate- and vigorous-intensity activity"

World Health Organisation

(In the USA, just 23% of adults achieve this!)

Interpreting Guidelines

Does it matter how it's spread across the week?

What is moderate?

What is vigorous?

"Adults should do at least 150 minutes of moderate-intensity aerobic physical activity throughout the week or do at least 75 minutes of vigorous-intensity aerobic physical activity throughout the week or an equivalent combination of moderate- and vigorous-intensity activity"

How are you supposed to combine?

Adding Metrics to Help

Wearables from Google include the notion of "Heart Points", and wearables from Garmin have "Intensity Minutes". These are essentially the same thing. They award:

1 point for every minute of moderate activity
2 points for every minute of vigorous activity

If you reach 150 points in a week, that is endorsed by the WHO as reaching the guideline minimum level.

So using sensing we apply the guideline to what the user does, and take the burden of tracking the activity volume away from the user.

Measuring Moderate and Vigorous

The intensity of an activity is ideally measured using energy expenditure measures. The gold standard is a VO_2 measurement but this is an invasive lab test.

Heart rate correlates well with intensity, as you might expect. So we can use HR to offload the decision of what is moderate and vigorous from the user.

(Aside: to some extent the guideline is "more movement is better", so even if our HR measurement is a bit wrong from PPG errors, it may still be valuable)

But...

Beware of Correlates...

Activity $\Rightarrow$ High HR
High HR $\not\Rightarrow$ Activity

Stress also elevates heart rate, but stress is bad (it results in Cortisol, prolonged exposure to which is very bad for your cardiovascular system).

So we need to add in activity recognition to heart rate measurement to decide whether to give heart points.

Lesson: When we switch from direct clinical signals to more convenient mobile signals, we must be careful to remember we often only have correlation and not (sole) causation.

Summary

Mobile health is a huge area.

We have touched on the (many) challenges inherent in just one part of it: mobile health sensing.

It's an exciting area where there is a real opportunity to make a meaningful, significant impact on the world.

<cough>
Great for a PhD...
</cough>